

# DNA Motifs Are Not General Predictors of Recombination in Two *Drosophila* Sister Species

James M. Howie<sup>1,†</sup>, Rupert Mazzucco<sup>1,†</sup>, Thomas Taus<sup>1,2,†</sup>, Viola Nolte<sup>1</sup>, and Christian Schlotterer<sup>1,\*</sup>

<sup>1</sup>Institut für Populationsgenetik, Vetmeduni Vienna, Austria

<sup>2</sup>Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Austria

\*Corresponding author: E-mail: christian.schlotterer@vetmeduni.ac.at.

<sup>†</sup>Equal first author.

Accepted: April 9, 2019

**Data deposition:** This project has been deposited at the European Nucleotide Archive under the accession numbers PRJEB29483 and ERP111789 and at Dryad, doi:10.5061/dryad.744p394.

## Abstract

Meiotic recombination is crucial for chromosomal segregation and facilitates the spread of beneficial and removal of deleterious mutations. Recombination rates frequently vary along chromosomes and *Drosophila melanogaster* exhibits a remarkable pattern. Recombination rates gradually decrease toward centromeres and telomeres, with a dramatic impact on levels of variation in natural populations. Two close sister species, *Drosophila simulans* and *Drosophila mauritiana* do not only have higher recombination rates but also exhibit a much more homogeneous recombination rate that only drops sharply very close to centromeres and telomeres. Because certain sequence motifs are associated with recombination rate variation in *D. melanogaster*, we tested whether the difference in recombination landscape between *D. melanogaster* and *D. simulans* can be explained by the genomic distribution of recombination rate-associated sequence motifs. We constructed the first high-resolution recombination map for *D. simulans* based on 189 haplotypes from a natural *D. simulans* population and searched for short sequence motifs linked with higher than average recombination in both sister species. We identified five consensus motifs significantly associated with higher than average chromosome-wide recombination rates in at least one species and present in both. Testing fine resolution associations between motif density and recombination, we found strong and positive associations genome-wide over a range of scales in *D. melanogaster*, while the results were equivocal in *D. simulans*. Despite the strong association in *D. melanogaster*, we did not find a decreasing density of these short-repeat motifs toward centromeres and telomeres. We conclude that the density of recombination-associated repeat motifs cannot explain the large-scale recombination landscape in *D. melanogaster*, nor the differences to *D. simulans*. The strong association seen for the sequence motifs in *D. melanogaster* likely reflects their impact influencing local differences in recombination rates along the genome.

**Key words:** *D. simulans*, genomic correlation, linkage disequilibrium, motif density, motif model, recombination map.

## Introduction

Meiotic recombination rate variation impacts on multiple important biological processes in sexual eukaryotes. It is crucial for chromosomal segregation (Roeder 1997; John 2005) but is also itself a powerful factor influencing genome organization and sequence variability (Aquadro et al. 1994; True et al. 1996). Meiotic recombination arises when a double-stranded break (DSB) leads to crossing over between homologous chromatids (Szostak et al. 1983; Schwacha and Kleckner 1995; Bergerat et al. 1997; Keeney et al. 1997; Hughes et al. 2018).

Higher rates of recombination break up genetic linkage and can increase the efficacy of natural selection (Haddrill et al. 2007; Charlesworth and Charlesworth 2010) and so affect the evolution of numerous genomic features. The reduction of transposable element density (Charlesworth et al. 1992, 1994; Rizzon et al. 2002; Petrov et al. 2011; Kofler et al. 2012) and the increased levels of DNA polymorphism in regions of high recombination (Begun and Aquadro 1992; Aquadro et al. 1994; Begun et al. 2007; Kulathinal et al. 2008) are probably the clearest examples.

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Yet although the eukaryotic meiotic machinery is generally highly conserved (Keeney 2001), rates of recombination have been observed to vary dramatically across species and populations, between individuals, and across sexes (Stapley et al. 2017; Haenel et al. 2018), apparently due to a combination of interacting environmental, epigenetic, and genetic factors (Detlefsen and Roberts 1921; Stern 1926; Neel 1941; Parsons 1958; Stapley et al. 2017). Moreover, the distribution of meiotic recombination rates among and along chromosomes varies markedly across taxa (Lichten and Goldman 1995; Petes 2001; Hey 2004; Choi and Henderson 2015; Hunter, Huang, et al. 2016; Stapley et al. 2017). Large-scale recombination suppression is often observed toward centromeres, the so-called “centromere effect” (Beadle 1932; Szauter 1984; Choulet et al. 2014; Hughes et al. 2018). Depending on the species, either suppression or enhancement of recombination has been observed toward the telomeres (Broman et al. 1998; Myers et al. 2005; Chan et al. 2012; Comeron et al. 2012). Heterochromatin, which is often associated with these regions, tends also to exhibit lower recombination rates than euchromatin (Sturtevant and Beadle 1936; Baker 1958; Roberts 1965; Szauter 1984; Termolino et al. 2016). Yet, in addition to these large-scale features of recombination landscapes, fast-evolving (Jeffreys et al. 2001) finer-scale variation can also be observed (Myers et al. 2005; Comeron et al. 2012).

It has been proposed that short sequence motifs are a key factor shaping the recombination landscape. For example, in humans a 13-mer, CCNCCNTNCCNC motif is targeted by the PRDM9 protein (Myers et al. 2010; Grey et al. 2011; Billings et al. 2013), via its zinc-finger array (Baudat et al. 2010; Parvanov et al. 2010), where it promotes histone methylation and meiotic crossover, reorganizing the nucleosome around it and driving DSB formation (Mihola et al. 2009; Brick et al. 2012; Baker et al. 2014; Pratto et al. 2014). These highly localized recombination events in 500–2,000-bp sections of chromosome have been called recombination “hotspots” (Lam and Keeney 2015). They are observed in a multitude of species including yeast, mice, humans among many others (Lam and Keeney 2015).

Hotspots are, however, no universal feature of recombination landscapes and are not observed in a range of species groups including *Caenorhabditis elegans* and *Drosophila* (Aquadro et al. 2001; Nachman 2002; Hey 2004; Chan et al. 2012; Manzano-Winkler et al. 2013; Heil et al. 2015; Miller et al. 2016). *Drosophila* spp. exhibit a large heterogeneity in recombination across their chromosomes, as demonstrated in *D. persimilis* (Stevison and Noor 2010), *D. pseudoobscura* (Cirulli et al. 2007; Kulathinal et al. 2008), and *D. melanogaster* (Singh et al. 2009; Comeron et al. 2012; Adrian et al. 2016). Still, *D. melanogaster* exhibits only a handful of mild “hotspots” relative to the ~30,000, often very strong hotspots observed in humans (International

HapMap Consortium 2007). Instead, the *D. melanogaster* recombination landscape is characterized by recombination “peaks” and “valleys” on a 5–500-kb scale (Singh et al. 2009; Chan et al. 2012; Comeron et al. 2012; Adrian et al. 2016), with which short “recombination motifs” are associated; as has also been seen in *D. pseudoobscura*, *D. persimilis*, and other *Drosophila* species (Cirulli et al. 2007; Kulathinal et al. 2008; Singh et al. 2009; Stevison and Noor 2010; Chan et al. 2012; Comeron et al. 2012; Heil and Noor 2012; Miller et al. 2012; Singh et al. 2013; Adrian et al. 2016). These repeat motifs, which often reside in transcription-associated euchromatic regions (Petes 2001; Comeron et al. 2012), are thought to increase the accessibility of DNA chromatin to double-stranded cleavage (Comeron et al. 2012) and destabilize DNA sequences, potentially in a stress, environmental or epigenetically dependent manner (Stern 1926; Neel 1941; Redfield 1966; Petes 2001; Hunter, Robinson, et al. 2016; Kohl and Singh 2018).

*Drosophila melanogaster*, *D. simulans*, and *D. mauritiana* are sister species which are ecologically and karyotypically similar (Lemeunier and Ashburner 1976; True et al. 1996) but differ dramatically in their recombination landscapes. Although *D. melanogaster* exhibits a characteristic gradual decrease in recombination rate toward centromeres and to a lesser extent also telomeres, the recombination landscape in *D. simulans* and *D. mauritiana* is much flatter with a rather constant recombination rate almost to the end of the chromosome arm, where it drops very quickly (True et al. 1996). Furthermore, these two species also have a higher recombination rate than *D. melanogaster* (True et al. 1996), which has been attributed, in *D. mauritiana*, to the MEI-218 protein which has highly diverged between *D. melanogaster* and *D. mauritiana*, and promotes recombination to a greater extent in the latter (Brand et al. 2018).

Here, to test the hypothesis that differences in genome-wide motif distributions can explain the observed differences in recombination rates (Adrian et al. 2016), we take a multi-step approach. First, we produce a high-resolution recombination map for *D. simulans* from 189 haplotypes. Next, we run a motif discovery in *D. melanogaster* and *D. simulans* and construct a consensus motif set for repeat motifs associated with recombination at the broadest scale, when each chromosome is divided into higher and lower than median recombination regions (each ~12,000 kb). We confirm the clear differences in recombination landscapes between the two species but find a similar set and distribution of recombination-associated motifs in each. Analyzing motif-recombination associations at a range of finer scales (1, 5, 25, 101, 501, and 2,501 kb), our results suggest that recombination-associated motifs cannot explain the large-scale differences in recombination landscapes between the two species but may have a significant impact on recombination at a local scale, in particular in *D. melanogaster*.

## Materials and Methods

### Recombination Map

#### Recombination Map Production

A total of 202 isofemale lines were established from a natural *D. simulans* population in Tallahassee, FL, USA, in 2010 (Barghi et al. 2017). From each of the 189 lines that were still alive in 2016, an individual male was selected and crossed with a virgin “reference” female from the M252 strain that was used to produce the *D. simulans* reference genome (Palmieri et al. 2015). DNA from individual F1 female offspring from each cross was then extracted using NEBNext Ultra II DNA Library Prep Kits (E7645, New England Biolabs). One hundred eighty-nine paired-end Illumina libraries were generated from single F1 females, with an average insert size of 260 bp, and sequenced on an Illumina HiSeq XTEN to obtain an average sequence coverage of 30×. Single-nucleotide polymorphisms (SNPs) were called with FreeBayes (v1.1.0-46-g8d2b3a0, Garrison and Marth 2012) relative to the *D. simulans* reference genome (Palmieri et al. 2015), requiring a minimum sequencing coverage of 10× and a variant quality of at least 50. All single-nucleotide polymorphisms that were polymorphic in the M252 reference strain were masked. Based on line-specific haplotype information, the genome-wide recombination map was estimated with LDJump (v0.1.4, Hermann et al. 2019), specifying a segment size of 1 kb, with an  $\alpha = 0.05$  and an  $\Theta = 0.04$ . We disabled LDJump’s segmentation analysis and worked with raw recombination rate estimates. Recombination rates were converted from  $\rho$  (rho) to units of cM/Mb by normalizing them so as to have a genetic map length between a set of marker genes equivalent to that which has been previously reported (True et al. 1996). The recombination map therefore represents an average of the recombination rate across both males and females in a population, normalized to females.

The resultant *D. simulans* recombination map was used in parallel with the *D. melanogaster* recombination map produced by Comeron et al. (2012), downloaded from the *D. melanogaster* Recombination Rate Calculator (Fiston-Lavier et al. 2010, last accessed April 19, 2019).

#### Recombination Map Scaling

As the raw recombination map output by LDJump is noisy, we smoothed each recombination map at several scales. In *D. melanogaster*, the raw map (Comeron et al. 2012) contained information on recombination rate at a 100-kb resolution, in *D. simulans* raw information was generated at a 1-kb scale. For smoothing, we used a moving median approach, using window sizes of 5, 25, 101, 501, and 2,501 kb for *D. simulans*, and of 101, 501, and 2,501 kb for *D. melanogaster*, respectively. It is important to note that identical smoothing was performed in both species, though for a larger number of finer resolutions in *D. simulans* due to the higher resolution of

the available data, such that species differences cannot arise due to different smoothing procedures. Advantages of the moving median as a smoothing method include low sensitivity to outliers, and a direct relationship to underlying data, in the sense that only values present in the raw data set can be present in the smoothed set if the median is taken based on an odd number of input values, which in our case it always was. The moving median is preferred to a moving mean due to a reduced sensitivity to outliers. Because this approach is computationally expensive and prone to deleting map features when there are long runs of identical values, we investigated, as an alternative approach, smoothing via LOESS local regression (Cleveland et al. 1992), which produces qualitatively equivalent results (supplementary fig. S2, Supplementary Material online). The smoothing scales chosen reflect those in Adrian et al. (2016), as relevant to potential motif explanatory power. The “correct” scale on which motifs may function is a priori unclear. A key point of our study is comparing results across several scales.

### DNA Motif Identification

#### Motif Discovery

For each species, we ran a genome-wide motif discovery using MEME (Bailey and Elkan 1994), from the MEME suite of motif-based sequence analysis tools (Bailey et al. 2009, version 5.0.1pl, accessible at <http://meme-suite.org> last accessed April 19, 2019; Bailey et al. 2015), a software designed to detect DNA sequence motifs in genetic data. After dividing each of the five large chromosomes (X, 2L, 2R, 3L, and 3R) into high- and low-recombining regions based on the chromosome median recombination rate, we used this software in the “differential enrichment” mode to detect motifs enriched in high-recombining areas of the genome. For *D. melanogaster*, we ran MEME on the release 5 reference genome (v. 5.36), for concordance with our recombination information from (Comeron et al. 2012). For *D. simulans*, we used the M252 Madagascar reference genome (Palmieri et al. 2015), to align with our recombination map. Motif discovery searches were run with species-specific Markov Background Models, simple matrices of background base frequencies obtained using the MEME *fasta-get-model* command, for each reference genome in turn. The full procedure was repeated with all smoothed maps (Recombination Map Production section). For completeness, a raw 1-kb window motif discovery run was also conducted for *D. simulans*. A similar search for motifs associated with lower recombination areas returned no results.

#### Motif Consensus Set

MEME motif discovery runs returned a set of 5, 4, and 3 motifs in *D. melanogaster* and 1, 2, 4, 1, 1, and 1 significant motifs in *D. simulans*, at the 101, 501, and 2,501, and

1, 5, 25, 101, 501, and 2,501-kb scales, respectively (supplementary S3, Supplementary Material online,  $E \leq 0.01$ ). That is, at least some motifs were recovered at all scales in both species, though fewer size scales were tested in *D. melanogaster* due to its coarser map resolution (see supplementary S3, Supplementary Material online, for a full description of all motifs recovered at all scales). It was noticed that, while individually distinct, numerous motifs contained similar core patterns while varying, for example, only in repeat number. As such, we constructed a set of five consensus motifs that captured the core variation in all motifs significantly associated with increased recombination, across both species, and over all scales. This core set of motifs C1–5 was determined via a two-step method. First, we contrasted the motifs across each of our recombination map smoothing scales in both species, retaining only motifs that occurred in at least one scale with a minimum significance of  $E \leq 0.01$  in at least one species. Motifs were then simplified by allowing only the most likely base at each position, and motif lengths were fixed as the longest sequence length that could be represented in both species (as lengths were by tendency longer in *D. melanogaster*). This resulted in the following set of consensus motifs: C1 = [A]<sub>11</sub>; C2 = [GCA]<sub>4</sub>; C3 = [CA]<sub>6</sub>; C4 = [TA]<sub>5</sub>; C5 = [G]<sub>8</sub>, C1–3 and C5 significant in both species, and C4 in *D. melanogaster* only (supplementary S3, Supplementary Material online). We note that *D. melanogaster* made the dominant contribution to the consensus motifs, as the motifs in *D. simulans* were less significant than those observed in *D. melanogaster* (supplementary S3, Supplementary Material online), and that the number of consensus motifs was informed by the data, and not decided a priori. As our consensus motifs turned out to be simplified versions of the most predictive motifs that were identified by Adrian et al. (2016), a subset of those sequence motifs identified in an earlier article by Comeron et al. (2012), we quantitatively tested and confirmed this similarity using the MEME Suite tool TomTom (Gupta et al. 2007), under default parameters (supplementary S4, Supplementary Material online).

## Genome-Wide Motif Densities

### Motif Locations

We converted the five consensus motifs into letter-probability matrices, to be used as input to Find Individual Motif Occurrences or FIMO, a MEME Suite tool designed to find genome-wide motif occurrences (Grant et al. 2011). Matrices were compiled in a hard, and a softer, version; with the expected base given a probability of 1 and unexpected bases probabilities of 0, or the expected base a probability of 0.97, and unexpected bases each a probability of 0.01. FIMO was then run for each species, taking the reference sequences and Markov Background Models as noted in the Motif Discovery section, and using parameter *max-stored-scores* = 50,000,000, and all others at default. Results of the

hard and soft motif probability runs were qualitatively identical, so hard coded motif probabilities were used for follow-up analysis (soft runs are not here reported).

### Motif Densities

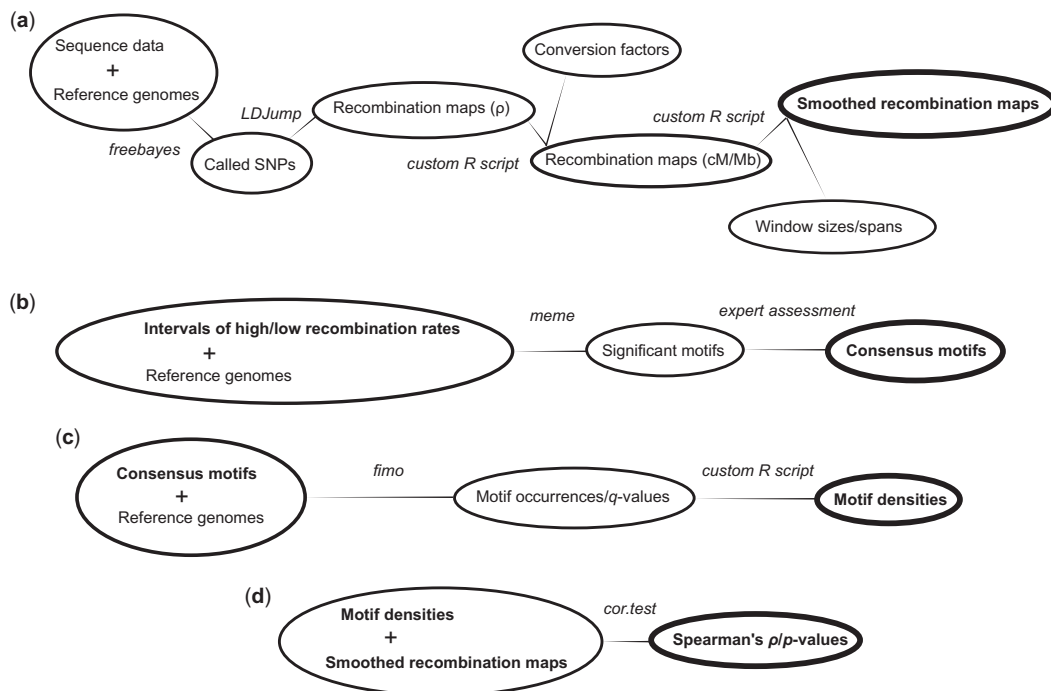
FIMO output provides, per motif, the genomic locations (chromosome, start and stop position) at which a motif was found, as well as a *P* value and a *q*-score (Benjamini and Hochberg 1995) per record, which show how well the motif was matched to the underlying reference sequence, respectively before and after correction for multiple testing (Benjamini and Hochberg 1995). To obtain genome-wide motif densities in each species, we calculated for each motif the sum of  $1 - q$ , across a sliding window of 1 kb, where *q* refers to the per record *q*-score, such that per window motif densities are discounted in relation to the quality of the motif match, with higher quality matches counting more. A total, genome-wide count (of  $1 - q$ ) of each motif was also obtained from the raw FIMO output. These motif densities were obtained for windows of 1, 5, 25, 101, 501, and 2,501-kb in *D. simulans* and at 101, 501, and 2,501 kb in *D. melanogaster*. We note that the scales at which the motifs were identified are far larger than the windows used for these motif density calculations.

## Motif–Recombination Correlations and Models

### Motif Density–Recombination Rate Correlations

To investigate the relationship between recombination rates and genome-wide abundances of individual motifs, we calculated the correlations between motif densities, binned at 1 kb, and corresponding recombination rates (cM/Mb), per motif, for *D. simulans* and *D. melanogaster*, respectively. As there was no clear a priori expectation for the genomic scale at which motifs would have most impact on recombination, the analysis was repeated for all three and five smoothing scales noted in the Recombination Map Scaling section for *D. melanogaster* and *D. simulans*, respectively (and was also repeated on the raw 1-kb scale in for *D. simulans*, not shown). The analysis was conducted using an identical procedure in each species, though at fewer scales in *D. melanogaster* due to limited resolution of the Comeron et al. (2012) recombination map. Spearman's rho,  $\rho$ , was used as a nonparametric estimator of the correlation between the test variables, and both the direction and significance of all correlations were extracted. To investigate the overall predictive power of motif densities, irrespective of chromosomal background, the analysis was repeated on the total genomic data, pooling across all of the five major chromosomes, with the analysis repeated per motif and species.

Finally, to test for explicit directional effects of each consensus motif on recombination, a linear regression model was fitted, per motif, species, scale, and chromosome, for the effect of motif density on local recombination rate, and repeated for the genome average.



**Fig. 1.**—A schematic representation of the bioinformatic pipeline used. Ovals represent physical data sets, lines represent tools used to derive them; see the Materials and Methods for details.

A schematic representation of this analytic pipeline is presented in figure 1. All statistical analyses were run in R, version 1.1383 (R Core Development Team 2018), using in-house scripts (fully available under [supplementary S5, Supplementary Material](#) online).

## Results

### Recombination Rates in *D. simulans* Are More Uniform across Chromosomes, than in *D. melanogaster*

We present the first high-resolution recombination map for *D. simulans*, and contrast it to that of *D. melanogaster* (Comeron et al. 2012). Across the range of smoothing parameters that we have applied to both species, the *D. simulans* recombination map is more uniform than that of *D. melanogaster* (figs. 2 and 3). The level of recombination suppression is lower toward the centromere in *D. simulans*. As in *D. melanogaster*, the main broad-scale features of the *D. simulans* map hold across the full range of genomic scales, whereas finer resolution peaks and troughs become visible only at higher resolutions, at the 5–501-kb scale (fig. 3). The finer-scale peaks (on a kb scale), as with the broader features (on a Mb scale), differ between these two sister species, and persist across smoothing scales (see figs. 2 and 3).

### Motif Density Landscapes Are Similar in *D. simulans* and *D. melanogaster*

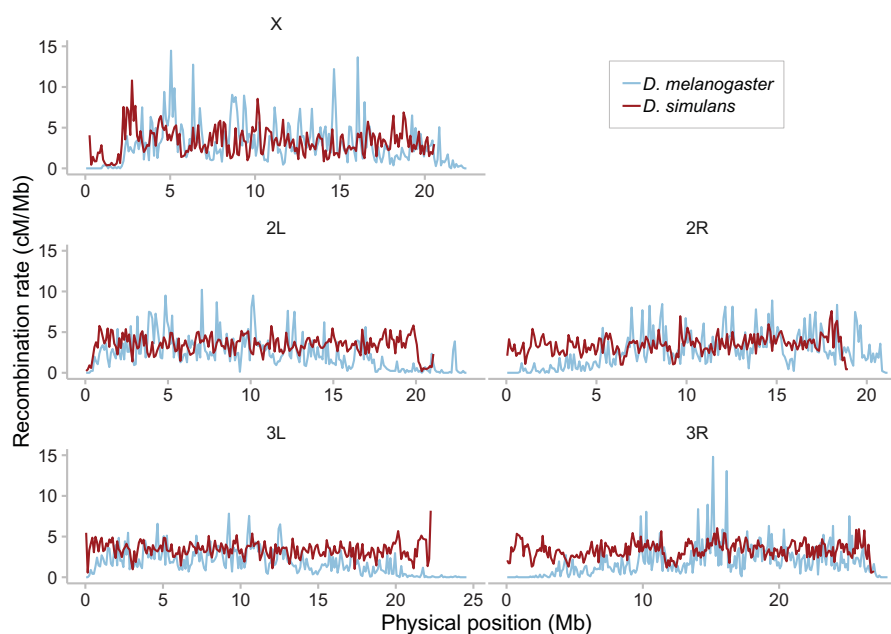
We identify five consensus motifs based on motifs recovered in each of the two species (Motif Consensus Set section) and

obtain their genome-wide densities. The consensus motifs were C1 = [A]<sub>11</sub>, C2 = [GCA]<sub>4</sub>, C3 = [CA]<sub>6</sub>, C4 = [TA]<sub>5</sub>, and C5 = [G]<sub>8</sub>, all five of which were significantly associated with recombination on at least one genomic resolution in *D. melanogaster* and present in both species, with C1–3 and C5 significantly associated at least once in *D. simulans*. Across all chromosomes and consensus motifs, the motif density landscapes were similar in *D. melanogaster* and *D. simulans* (fig. 4). This was especially true for intermediate size landscape features, such as humps and wider valleys (e.g., motif C2 on X, 7.5-Mb position, or 2L at the 8 and 12-Mb positions, fig. 4). Therefore, motif density cannot explain the differences in the broad recombination landscape between both species.

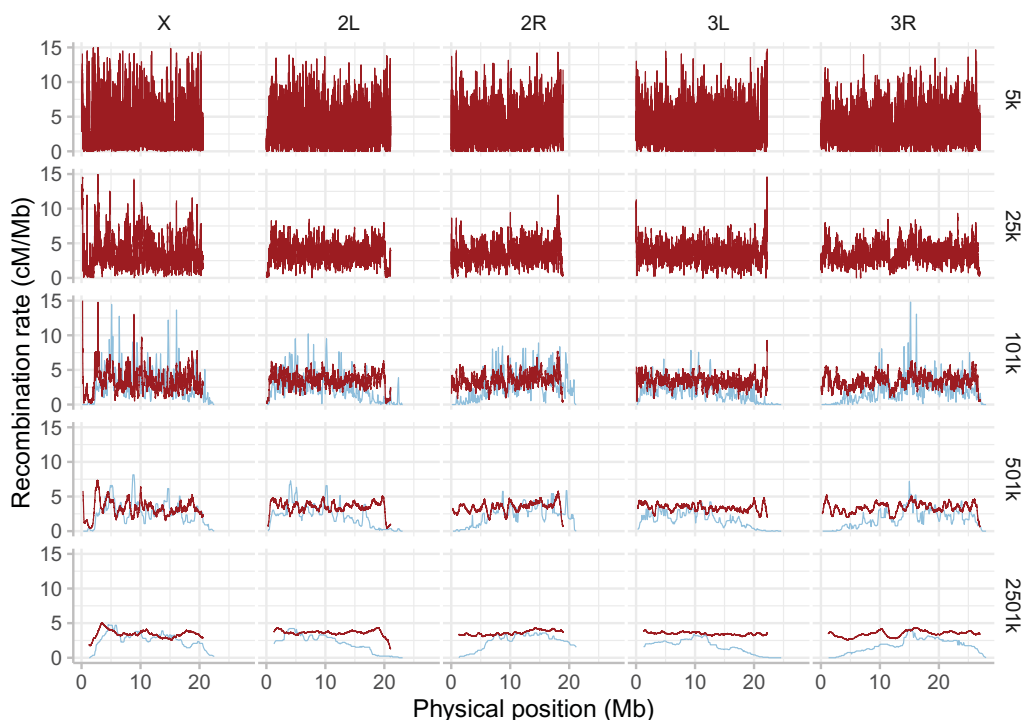
Finer resolution peaks and troughs varied more between species (e.g., motif C4 on X, 5–15-Mb position, fig. 4). Further, although the different motifs, C1–5, displayed similar broad patterns in each species—per chromosome and genome-wide—some species-specific patterns were seen. Motifs C1, [A]<sub>11</sub> and C5, [G]<sub>8</sub> were far less common in *D. simulans*, which had a lower total motif count, whereas the opposite was true for motif C4, [TA]<sub>5</sub>. Nonetheless, genome-wide motif distributions were similar in each species.

### Associations between Motif Densities and Recombination Rates Are Generally Weaker and Less Significant in *D. simulans* than in *D. melanogaster*

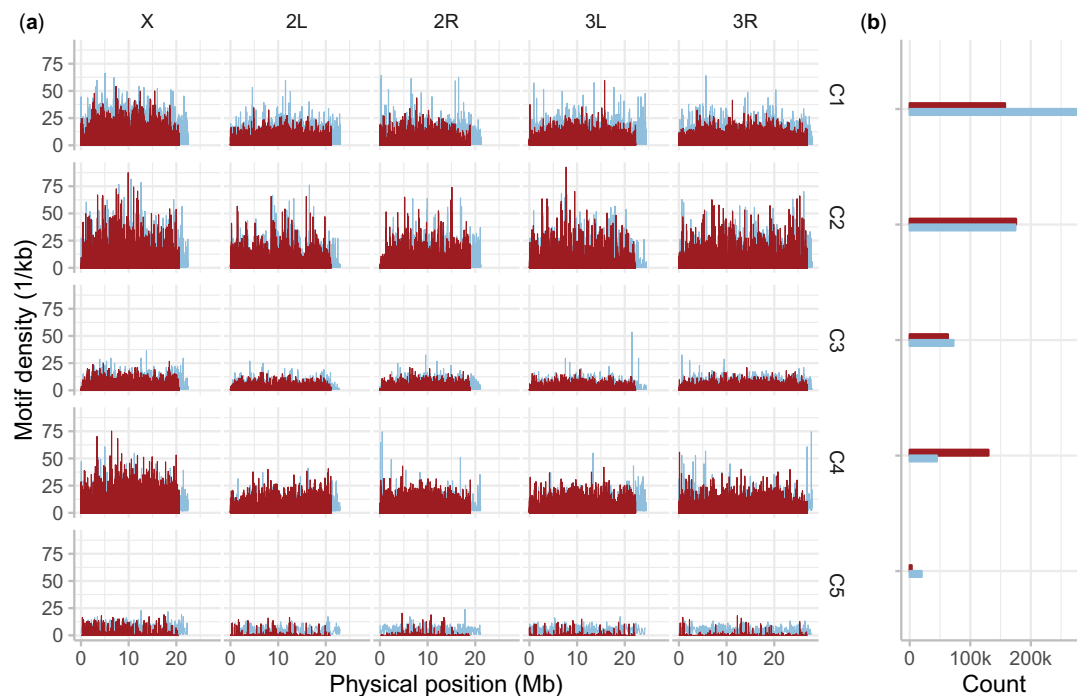
We examined correlations between motif densities and recombination rates in each species, both per chromosome,



**FIG. 2.**—Recombination rates in *Drosophila simulans* are more uniform across chromosomes than in *Drosophila melanogaster*. Red lines show the recombination rate in *D. simulans* for each of the major chromosomes (name labels in top margin), smoothed in this case at a 101-kb window size with a moving median. For comparison, blue lines show the recombination rate in *D. melanogaster* (with data taken from Cameron et al. [2012]); figure 3 for other resolutions.



**FIG. 3.**—Recombination rates in *Drosophila simulans* are more uniform across chromosomes than in *Drosophila melanogaster*, at all smoothing scales. Red lines show the recombination rate in *D. simulans* for each of the major chromosomes (names in top margin), smoothed now at five different window sizes (right margin, in bp) with a moving median. For comparison, blue lines show the recombination rate in *D. melanogaster* (data taken from Cameron et al. [2012] at 101 kb; and smoothed at 501 and 2,501 kb; with data not available at smaller resolutions).



**FIG. 4.**—Motif densities are similar in *Drosophila simulans* and *Drosophila melanogaster*. (a) Red (*D. simulans*) and blue (*D. melanogaster*) lines show motif densities across major chromosomes (top margin) as reported by FIMO, with motif occurrences discounted by  $1 - q$  (see main text) and binned into 1-kb windows for each consensus motif (C1–5). (b) Total motif counts across all five large chromosomes. FIMO threshold:  $P$  value of  $1e-4$  (default setting).

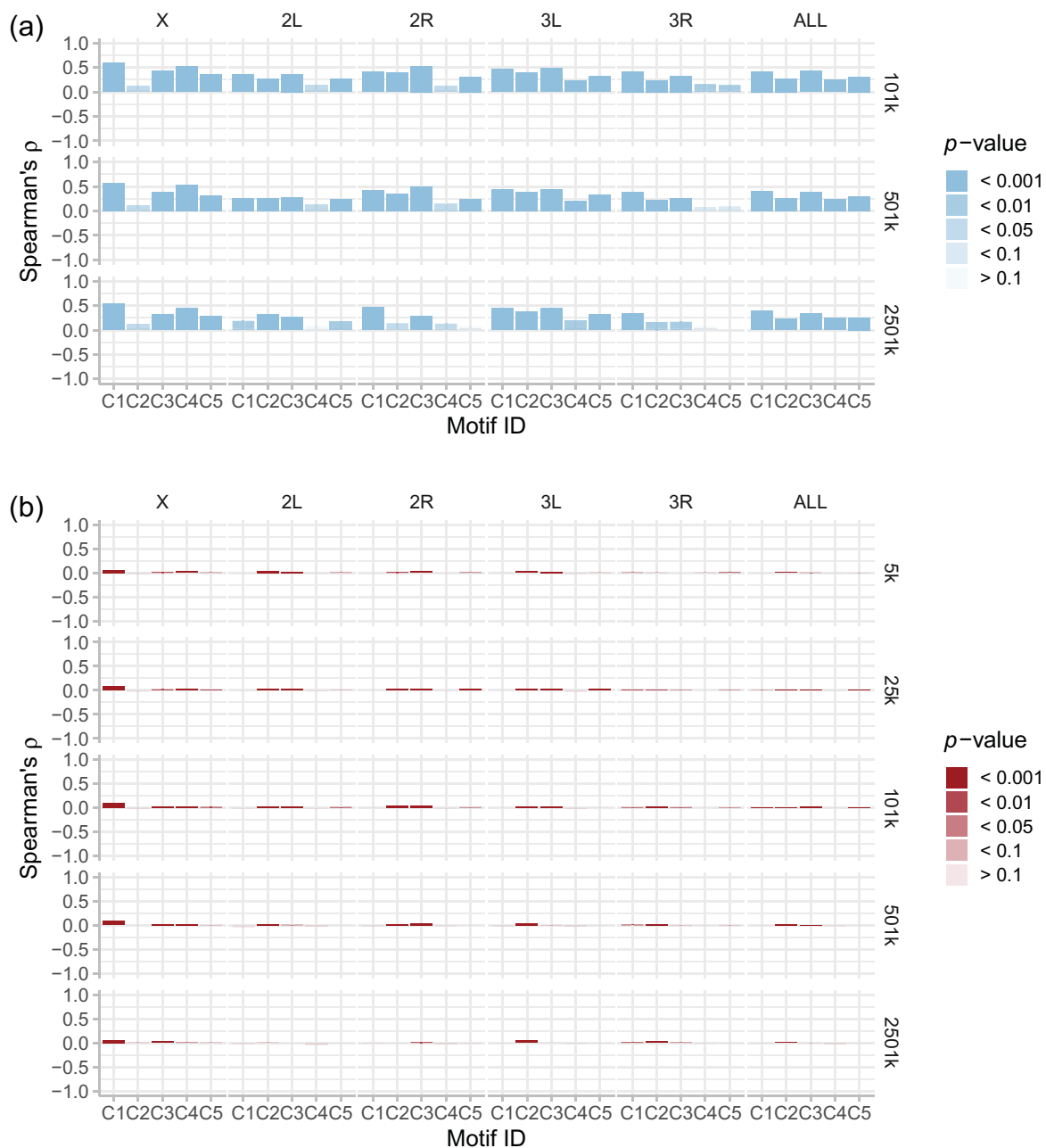
and genome-wide, and at a range of genomic scales. A clear difference was observed between the species. In *D. melanogaster*, all but one correlation was positive, most were highly significant both genome-wide and per chromosome, and the correlation coefficients (Spearman's  $\rho$ ) were generally large; with a range of  $\sim 0.4$ – $0.6$  for the most associated motifs per chromosome (and genome-wide, fig. 5a). In contrast, the associations observed in *D. simulans* were heterogeneously positive or negative, had lower significances than those observed in *D. melanogaster*, and were in all cases weak; with a range of  $\sim 0.01$ – $0.04$  for the most associated motifs per chromosome (and genome-wide, fig. 5b). In both species, there was also variation in the importance of different motifs on different chromosomes (see below). However, although in *D. melanogaster*, the patterns of motif association held across all scales for each chromosome and genome-wide, in *D. simulans* there were occasional exceptions to this rule. For instance, on 2L, 2R, 3L, and genome-wide, the positive correlations for C1 and C4 switched direction at scales larger than 25–101 kb. Given that these correlations were very weak with low significance, we attribute these discrepancies stochastic noise, rather than biological signals. We finally note that motifs C1, C2, and C3 were the most associated with recombination across most major chromosomes in both species (though to a far lesser extent in *D. simulans*), but that an exception is observed for the X chromosome. Here, motif C2 had a very weak association with

recombination rate in both species, and motif C4 instead had a high association, relative to its weak association on most autosomes in both species. Very similar observations were seen for the linear regressions (supplementary fig. S1, Supplementary Material online), with more models being significant and positive for *D. melanogaster*.

## Discussion

We present the first high-resolution recombination map for *D. simulans*, and a comparative analysis of recombination-associated motifs and their association with recombination in two sister species, *D. melanogaster* and *D. simulans*. We tested the hypothesis that such motifs predict recombination rates within the *D. melanogaster* species subgroup.

Our *D. simulans* recombination map confirms the results of previous, lower resolution work in this species (Stuktevanat 1929; Ohnishi and Voelker 1979, 1981; True et al. 1996). We find that the *D. simulans* recombination landscape is far flatter than in *D. melanogaster* (figs. 2 and 3). Although centromeric recombination suppression on the X, and to some extent on 2L and 3R, is observed in *D. simulans*, it is restricted to a small genomic region, whereas in *D. melanogaster* the recombination rate decreases only gradually over a much larger region in proximity to the centromeres (Comeron et al. 2012). In *D. simulans*, a similarly sharp telomeric suppression is observed on 2L (and to some extent on X, 2R, 3L, and 3R) at most



**FIG. 5.**—Associations between motif densities and recombination rates are generally weaker and less significant in *Drosophila simulans* than in *Drosophila melanogaster*. For (a) *D. melanogaster*, and (b) *D. simulans*, bars indicate Spearman's rho,  $\rho$ , (height) and the corresponding  $P$  value (transparency), from tests of the correlation between motif densities (as shown in fig. 4 for *D. simulans*, but rebinned for *D. melanogaster* to account for the resolution of the available data) and recombination rates, across individual chromosomes and for the five large chromosomes together (top margin), at all smoothing levels (see right margin).

smoothing scales (this pattern is less clear at 2,501 kb). Unlike in *D. melanogaster* (Comeron et al. 2012), overall recombination rates in *D. simulans* appear quite similar between X and the autosomes (figs. 2 and 3).

We caution however, that recombination rate estimates from population polymorphism data are sensitive to demographic events, and particularly the ratio of X-chromosomal

and autosomal variation differs widely between populations (Kauer et al. 2002; Schöfl and Schlötterer 2004). In *D. simulans* and *D. melanogaster*, mid-to-large-scale recombination features clearly persist over the 101, 501, and 2,501-kb smoothed maps. In *D. simulans*, our high-resolution map shows that such features also persist down to the 25 and 5-kb scale (e.g., the dip on 3R at 12.5-Mb position, fig. 3).



As with the centromeric differences however, mid-scale and narrower landscape features differ between the species, especially at the 101 and 501-kb resolutions. In short, at all genomic scales tested the two species differ dramatically in recombination rates, over broad- and finer-scale recombination features.

One important difference between the recombination maps between the two species is that for *D. simulans*, we inferred the recombination landscape from population polymorphism data, whereas the recombination map of *D. melanogaster* is based on recombination events occurring in the laboratory. Because polymorphism patterns in natural populations are affected by selection and demography, this may impact the inferred recombination map. In particular the well-studied influence of recombination rate on the patterns of natural variation in positive (hitchhiking, Maynard Smith 1974; Hartfield and Otto 2011) and negative (background selection, Charlesworth et al. 1995; Comeron 2017) selection will also affect any polymorphism-based recombination maps. With regards to the inferred *D. simulans* recombination map, it is important to note that we inferred a rather similar recombination rate along the chromosomes, strongly suggesting very limited differences in  $N_e$ , such that while our recombination map based on  $\rho$  will incorporate information about  $N_e$  ( $\rho = N_e \times r$ ) the minimal effects of variable linked selection along the chromosomes mean this is unlikely to produce biases between the maps used for *D. melanogaster* and *D. simulans* in this study. Hence, we consider population data based recombination map inference a reliable method, despite its obvious limitations.

Direct implications from these differences in genetic maps are that linkage disequilibrium should be both lower and less variable across the *D. simulans* chromosomes relative to those of *D. melanogaster*. It is important to keep in mind though that the *D. simulans* reference genome includes less repetitive DNA at the centromeric and telomeric ends of the chromosomes, so a comparison of recombination rates is not possible at the extremes of these regions. Nonetheless, our results bolster the current understanding of *D. simulans* recombination as less heterogeneous than that of *D. melanogaster* (True et al. 1996; Comeron et al. 2012), and indicate that selection will be generally more efficient in *D. simulans*, as genes that are uncoupled by recombination may result in more distinct signals of selection, in particular in Evolve and Resequencing experiments (Kofler and Schlötterer 2014; Tobler et al. 2014; Barghi et al. 2017). Hence, adaptive evolutionary changes may occur more rapidly in *D. simulans*, all else being equal, because Hill-Robertson effects are reduced by the higher recombination (Hill and Robertson 1966).

Turning to the causes of this recombination variation, we ran a MEME motif search to identify short DNA sequence motifs associated with regions of higher than average recombination at a very coarse scale, repeating this search in both *D. melanogaster*, and *D. simulans*. The first point of note was

that a larger number of motifs were returned in *D. melanogaster*, and that those in *D. simulans* were by tendency both shorter and showed a less significant association with recombination rate, with lower quality matches. Nonetheless, a generally similar set of motifs was recovered in each species, and across each recombination map smoothing scale investigated. In short, we obtained a subset of the *D. melanogaster* motifs in *D. simulans*; motifs C1, C2, C3, and C5, providing some confidence in the impact of these motifs on the recombination rate. The motif sharing between the two *Drosophila* species also provides some evidence that recombination motifs may to some degree be universal across *Drosophila* species. This idea builds upon prior work, which has shown that there is some overlap in motifs between more distant *Drosophila* species, such as *D. pseudoobscura*, which exhibits CACAC (Cirulli et al. 2007), CCCCACCCC, and CCTCCCT motifs (Kulathinal et al. 2008), and *D. persimilis*, which exhibits a CCNCCNTNNCCNC motif (Stevison and Noor 2010). This led Comeron et al. (2012) to speculate that *Drosophila* has a stable set of recombination motifs of universal function, which they confirmed in part by showing that *D. melanogaster* also exhibit the CACAC and CCTCCCT motifs, though not the CCCCACCCC motif. Our study builds on this result, showing that a larger degree of motif overlap can be seen both when contrasting consensus motifs and when comparing between more closely related species, and that the  $[CA]_n$  motif is universal to all *Drosophila* species studied. However, it is immediately notable that no complex, multipart motifs like CCNCCNTNNCCNC were recovered in our study.

The genome-wide distribution of motifs (fig. 4) revealed, somewhat surprisingly, that there are also clear parallels between the two species motif landscapes. Not only do motifs with higher density in *D. melanogaster* generally have a higher density in *D. simulans*, but the patterns of motif distribution genome-wide are also remarkably similar. For instance, a similar “hump” and “peak” can be observed at the 8 and 9-Mb positions of chromosomes X and 2L, respectively, for motif C2, in both species, whereas a density “trough” can be seen at 15 Mb on chromosome 2L for this motif (fig. 4). Motifs C1, C3, and C4 likewise exhibit very limited differences between species, on all chromosomes (fig. 4), despite clear differences in recombination rates (fig. 3). A few differences do exist. Motif C1 is more common in *D. melanogaster*, even if the “landscape” is similar to *D. simulans*; Motif C5 is then less common in *D. simulans*, and exhibits a distinct landscape on all autosomes, whereas any narrow-scale features rarely overlap between species, mirroring patterns of distinct recombination peaks and similar landscapes seen in *D. melanogaster* populations (Chan et al. 2012; Heil et al. 2015). Consequently, although it might be tempting to speculate that subtle differences in motif densities can explain the flatter recombination landscape of *D. simulans* and its unique recombination peak set, it is difficult to reconcile the distinctive

patterns of recombination rate variation in the two species with their exceptionally similar motif density landscapes, that are almost identical between species, especially when focusing on the large-scale recombination differences that we observed in the centromeric and telomeric regions.

The similar motif density patterns between the two species cast doubt on the hypothesis that differences in motif distribution can account for differences in recombination variation in these species. If divergent motif densities really account for the species differences in recombination rates, how can we explain the discordance seen in the reduced recombination toward the centromeres in *D. melanogaster*, the lack of this reduction in *D. simulans*, and the similar motif distributions over these regions in both species? To investigate this observation quantitatively, we calculated Spearman's rho,  $\rho$ , as an estimator of the correlation between genome-wide motif density and recombination rate (cM/Mb), for each motif, in each species, across a range of smoothing scales. This revealed a striking difference between the two species. In *D. melanogaster*, all associations (aside one) were positive, for all motifs at all scales tested, with low *P* values observed in most cases (fig. 5). These results accord well with those of Adrian et al. (2016), who found positive associations between motif densities and recombination rate in *D. melanogaster*, using a similar set of motifs (see [supplementary S4, Supplementary Material](#) online). In contrast, the associations observed in *D. simulans* were far smaller, and far more heterogeneous across chromosomes and motifs (fig. 5). This observation was confirmed by our linear regression models, fitted to explicitly test the predictive power of each motif to explain recombination rate variation, which showed an almost identical pattern to these correlations ([supplementary fig. S1, Supplementary Material](#) online). The correlational and model fit patterns were similar within each species across all smoothing scales, which is to say that the strength of the correlations observed did not increase or decrease with the higher or lower resolution of the recombination maps, at different scales. The clear implication is that motif densities do not universally predict recombination rates across the *Drosophila* clade, and are in particular not responsible for the large-scale differences observed between our two species. It is therefore pertinent to ask what alternative mechanisms could explain such differences.

A strong candidate is the dicistronic meiosis gene *mei-217/mei-218* and its protein product, MEI-218 (Brand et al. 2018), which is involved in the resolution of crossing over into DSBs and recombination (Brand et al. 2018). Divergent forms have recently been identified in *D. mauritiana* and *D. melanogaster*, species that diverged 0.6–0.9 Ma. Like *D. simulans*, *D. mauritiana* exhibits a higher and flatter recombination rate landscape than *D. melanogaster* (True et al. 1996), with the difference especially pronounced in the centromeric and telomeric regions (True et al. 1996), and with this pattern expressed to an even larger extent than is seen in *D. simulans*

(True et al. 1996). Intriguingly then, Brand et al. (2018) also found a high divergence in DNA and protein structure in the *mei-217/mei-218* gene and MEI-218 protein between *D. mauritiana* and *D. melanogaster*. The *D. mauritiana* form was far more effective in promoting recombination, increasing recombination assurance and reducing crossover interference (Brand et al. 2018). It explained a large portion of the variance in crossover rates between *D. mauritiana* and *D. melanogaster*, especially that in the centromeric and telomeric regions (Brand et al. 2018), and so could be a primary mechanistic variant explaining the differences in recombination between *D. simulans* and *D. melanogaster*. The clear parallel differences between the recombination maps of *D. melanogaster* versus *D. mauritiana* and *D. melanogaster* versus *D. simulans* imply that *mei-217/mei-218* may be responsible for the heterogeneity in recombination landscape that we have observed, a possibility that should be explored and tested in future work.

What then might explain the clear correlations between motif density and recombination seen here in *D. melanogaster*, but not *D. simulans*? A simple explanation is that motifs are responsible for variation in recombination rate on a local scale. Hence, the lower density in *D. simulans* results also in less microscale variation in recombination rate. Alternatively, this pattern could be explained if the recombination motifs are recognized directly by cleavage proteins, similar to PRDM9 that differ in function or effectiveness between *D. simulans* and *D. melanogaster*. Recent evidence shows that a zinc-finger gene and protein of this type exists in *D. melanogaster* (Hunter, Huang, et al. 2016). Yet, such proteins tend to bind to complex, rather than short-repeat motifs, making this explanation unlikely. Another possibility relates to chromatin structure, because short-repeat DNA recombination motifs are thought to play roles in loosening chromatin structure, increasing access for DSB inducing proteins (Comeron et al. 2012; Adrian et al. 2016, and references therein). This could account for microvariation in recombination rates genome-wide between species, for instance because the motifs were generally shorter and so presumably less effective at chromatin loosening in *D. simulans*, genome-wide. Circumstantial evidence in favor of this hypothesis includes that in both species motif correlation patterns varied cross chromosomes – for instance, C4 was a good predictor only on X—suggesting that motifs can operate in a context dependent manner. Likewise, the removal of subcentromeric and subtelomeric region recombination data has been found not to alter correlational patterns in *D. melanogaster* (Adrian et al. 2016), suggesting that if motifs densities explain some recombination rate genome wide, they cannot explain centromeric and telomeric differences.

In short, we present the hypothesis that although short-repeat DNA motifs may affect recombination at a microscale, genome-wide, for instance in relation to euchromatic structure context, they cannot explain the large differences in

recombination landscape differences between species, especially in the centromeric and telomeric regions. This variation seems far more likely to be explained by a mechanism such as *mei-217/mei-218*.

## Data Accessibility

Raw sequence reads for the 189 isofemale line haplotypes are available to download at the European Nucleotide Archive (ENA) under the primary and secondary accession numbers PRJEB29483 and ERP111789. Phased haplotypes are available from Dryad (doi:10.5061/dryad.744p394). Finally, CSV and MimicrEE ready text files for the *D. simulans* and *D. melanogaster* recombination maps are available at all resolutions from Dryad (doi:10.5061/dryad.744p394), as are the raw recombination rates output from LDJump in rho,  $\rho$ , before conversion to a cM/Mb scale (doi:10.5061/dryad.744p394), as well as motif density files, for genomic position and discounted motif count score (doi:10.5061/dryad.744p394).

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

The authors thank all members of the Institute of Population Genetics for discussion and support on this project, as well as two anonymous reviewers for their comments. This work was supported by the European Research Council (ERC) grant “ArchAdapt” to C.S., an Austrian Academy of Sciences DOC fellowship to T.T., and the Austrian Science Fund (FWF, W1225).

## Author Contributions

R.M., J.M.H., and C.S. conceived the study and interpreted the results. T.T. produced the recombination map. V.N. generated the NGS libraries. R.M. performed the analysis, with input from J.M.H. J.M.H. and C.S. wrote the manuscript with input from all authors.

## Literature Cited

- Adrian AB, Corchado JC, Comeron JM. 2016. Predictive models of recombination rate variation across the *Drosophila melanogaster* genome. *Genome Biol Evol.* 8(8):2597–2612.
- Ashburner M, Lemeunier F. 1976. Relationships within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*). I. Inversion polymorphisms in *Drosophila melanogaster* and *Drosophila simulans*. *Proc R Soc Lond B Biol Sci.* 193(1111):137–57.
- Aquadro CF, Bauer DuMont V, Reed FA. 2001. Genome-wide variation in the human and fruitfly: a comparison. *Curr Opin Genet Dev.* 11(6):627–634.
- Aquadro CF, Begun DJ, Kindahl EC. 1994. Selection, recombination, and DNA polymorphism in *Drosophila*. In: *Non-neutral evolution*. Boston, MA:Springer. p. 46–56.
- Bailey TL, Elkan C, editors. 1994. *Proceedings of the second international conference on intelligent systems for molecular biology*. Menlo Park (CA):The AAAI Press.
- Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Res.* 43(W1):W39–W49.
- Bailey TL, et al. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37(Web Server):W202–W208.
- Baker CL, Walker M, Kajita S, Petkov PM, Paigen K. 2014. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Res.* 24(5):724–732.
- Baker WK. 1958. Crossing over in heterochromatin. *Am Nat.* 92(862):59–60.
- Barghi N, Tobler R, Nolte V, Schlötterer C. 2017. *Drosophila simulans*: a species with improved resolution in evolve and resequence studies. *G3 (Bethesda)* 7:2337–2343.
- Baudat F, et al. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327(5967):836–840.
- Beadle GW. 1932. A possible influence of the spindle fibre on crossing-over in *Drosophila*. *Proc Natl Acad Sci U S A.* 18(2):160–165.
- Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356(6369):519–520.
- Begun DJ, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5(11):e310.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 57:289–300.
- Bergerat A, et al. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* 386(6623):414–417.
- Billings T, et al. 2013. DNA binding specificities of the long zinc-finger recombination protein PRDM9. *Genome Biol.* 14(4):R35.
- Brand CL, Cattani MV, Kingan SB, Landeen EL, Presgraves DC. 2018. Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. *Curr Biol.* 28(8):1289–1295.e1284.
- Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485(7400):642–645.
- Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. 1998. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet.* 63(3):861–869.
- Chan AH, Jenkins PA, Song YS. 2012. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLoS Genet.* 8(12):e1003090.
- Charlesworth B, Charlesworth D. 2010. *Elements of evolutionary genetics*. Englewood (CO): Roberts and Company Publishers.
- Charlesworth B, Lapid A, Canada D. 1992. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genet Res.* 60(2):103–114.
- Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371(6494):215–220.
- Charlesworth D, Charlesworth B, Morgan MT. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics* 141(4):1619–1632.
- Choi K, Henderson IR. 2015. Meiotic recombination hotspots—a comparative view. *Plant J.* 83(1):52–61.

- Choulet F, et al. 2014. Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345(6194):1249721.
- Cirulli ET, Kliman RM, Noor MA. 2007. Fine-scale crossover rate heterogeneity in *Drosophila pseudoobscura*. *J Mol Evol*. 64(1):129–135.
- Cleveland WS, Grosse E, Shyu WM. 1992. Local regression models. In: Chambers JM, Hastie TJ, editors. *Statistical models in S*. Wadsworth & Brooks/Cole.
- Comeron JM. 2017. Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. *Philos Trans R Soc Lond B Biol Sci*. 372(1736). pii: 20160471.
- Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet*. 8(10):e1002905.
- Detlefsen JA, Roberts E. 1921. Studies on crossing over—I. The effect of selection on crossover values. *J Exp Zool*. 32:333–354.
- Fiston-Lavier AS, Singh ND, Lipatov M, Petrov DA. 2010. *Drosophila melanogaster* Recombination Rate Calculator. *Gene* 463(1-2):18–20.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv: 1207.3907.
- Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017–1018.
- Grey C, et al. 2011. Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol*. 9(10):e1001176.
- Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. *Genome Biol*. 8(2):R24.
- Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biol*. 8(2):R18.
- Haenel Q, Laurentino TG, Roesti M, Berner D. 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Mol Ecol*. 27(11):2477–2497.
- Hartfield M, Otto SP. 2011. Recombination and hitchhiking of deleterious alleles. *Evolution* 65(9):2421–2434.
- Heil CS, Ellison C, Dubin M, Noor MA. 2015. Recombining without hotspots: a comprehensive evolutionary portrait of recombination in two closely related species of *Drosophila*. *Genome Biol Evol*. 7:2829–2842.
- Heil CS, Noor MA. 2012. Zinc finger binding motifs do not explain recombination rate variation within or between species of *Drosophila*. *PLoS One* 7(9):e45055.
- Hermann P, Heissl A, Tiemann-Boege I, Futschik A. 2019. LDJump: estimating variable recombination rates from population genetic data. *Mol Ecol Resour*. 1–16.
- Hey J. 2004. What's so hot about recombination hotspots? *PLoS Biol*. 2(6):e190.
- Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res*. 8(3):269–294.
- Hughes SE, Miller DE, Miller AL, Hawley RS. 2018. Female meiosis: snags, recombination, and segregation in *Drosophila melanogaster*. *Genetics* 208(3):875–908.
- Hunter CM, Huang W, Mackay TF, Singh ND. 2016. The genetic architecture of natural variation in recombination rate in *Drosophila melanogaster*. *PLoS Genet*. 12(4):e1005951.
- Hunter CM, Robinson MC, Aylor DL, Singh ND. 2016. Genetic background, maternal age, and interaction effects mediate rates of crossing over in *Drosophila melanogaster* females. *G3 (Bethesda)* 6:1409–1416.
- International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 449(7164):851–861.
- Jeffreys AJ, Kauppi L, Neumann R. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet*. 29(2):217–222.
- John B. 2005. *Meiosis*. New York: Cambridge University Press.
- Kauer M, Zangerl B, Dieringer D, Schlötterer C. 2002. Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. *Genetics* 160(1):247–256.
- Keeney S. 2001. Mechanism and control of meiotic recombination initiation. *Curr Top Dev Biol*. 52:1–53.
- Keeney S, Giroux CN, Kleckner N. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88(3):375–384.
- Kofler R, Betancourt AJ, Schlötterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet*. 8(1):e1002487.
- Kofler R, Schlötterer C. 2014. A guide for the design of evolve and resequencing studies. *Mol Biol Evol*. 31(2):474–483.
- Kohl KP, Singh ND. 2018. Experimental evolution across different thermal regimes yields genetic divergence in recombination fraction but no divergence in temperature associated plastic recombination. *Evolution* 72(4):989–999.
- Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MA. 2008. Fine-scale mapping of recombination rate in *Drosophila* refines its correlation to diversity and divergence. *Proc Natl Acad Sci U S A*. 105(29):10051–10056.
- Lam I, Keeney S. 2015. Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harb Perspect Biol*. 7(1):a016634.
- Lichten M, Goldman AS. 1995. Meiotic recombination hotspots. *Annu Rev Genet*. 29(1):423–444.
- Manzano-Winkler B, McGaugh SE, Noor MA. 2013. How hot are *Drosophila* hotspots? Examining recombination rate variation and associations with nucleotide diversity, divergence, and maternal age in *Drosophila pseudoobscura*. *PLoS One* 8(8):e71582.
- Maynard Smith J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res*. 23:35.
- Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323(5912):373–375.
- Miller DE, et al. 2012. A whole-chromosome analysis of meiotic recombination in *Drosophila melanogaster*. *G3 (Bethesda)* 2:249–260.
- Miller DE, et al. 2016. Whole-genome analysis of individual meiotic events in *Drosophila melanogaster* reveals that noncrossover gene conversions are insensitive to interference and the centromere effect. *Genetics* 203(1):159–171.
- Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310(5746):321–324.
- Myers S, et al. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* 327(5967):876–879.
- Nachman MW. 2002. Variation in recombination rate across the genome: evidence and implications. *Curr Opin Genet Dev*. 12(6):657–663.
- Neel JV. 1941. A relation between larval nutrition and the frequency of crossing over in the third chromosome of *Drosophila melanogaster*. *Genetics* 26(5):506–516.
- Ohnishi S, Voelker RA. 1979. Comparative studies of allozyme loci in *Drosophila simulans* and *D. melanogaster*. II. Gene arrangement on the third chromosome. *Jpn J Genet*. 54:203–209.
- Ohnishi S, Voelker RA. 1981. Comparative studies of allozyme loci in *Drosophila simulans* and *Drosophila melanogaster*. I. Three dipeptidase loci. *Biochem Genet*. 19(1-2):75–85.
- Palmieri N, Nolte V, Chen J, Schlötterer C. 2015. Genome assembly and annotation of a *Drosophila simulans* strain from Madagascar. *Mol Ecol Resour*. 15(2):372–381.
- Parsons PA. 1958. Selection for increased recombination in *Drosophila melanogaster*. *Am Nat*. 92(865):255–256.

- Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. *Science* 327(5967):835.
- Petes TD. 2001. Meiotic recombination hot spots and cold spots. *Nat Rev Genet.* 2(5):360–369.
- Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, Gonzalez J. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol.* 28(5):1633–1644.
- Pratto F, et al. 2014. DNA recombination. Recombination initiation maps of individual human genomes. *Science* 346(6211):1256442.
- R Core Development Team. 2018. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
- Redfield H. 1966. Delayed mating and the relationship of recombination to maternal age in *Drosophila melanogaster*. *Genetics* 53(3):593–607.
- Rizzon C, Marais G, Gouy M, Biemont C. 2002. Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Res.* 12(3):400–407.
- Roberts PA. 1965. Difference in the behaviour of eu- and heterochromatin: crossing-over. *Nature* 205:725–726.
- Roeder GS. 1997. Meiotic chromosomes: it takes two to tango. *Genes Dev.* 11(20):2600–2621.
- Schöfl G, Schlötterer C. 2004. Patterns of microsatellite variability among X chromosomes and autosomes indicate a high frequency of beneficial mutations in non-African *D. simulans*. *J Mol Biol Evol.* 21(7):1384–1390.
- Schwacha A, Kleckner N. 1995. Identification of double Holliday junctions as intermediates in meiotic recombination. *Cell* 83(5):783–791.
- Singh ND, Aquadro CF, Clark AG. 2009. Estimation of fine-scale recombination intensity variation in the white-echinus interval of *D. melanogaster*. *J Mol Evol.* 69(1):42–53.
- Singh ND, Stone EA, Aquadro CF, Clark AG. 2013. Fine-scale heterogeneity in crossover rate in the garnet-scalloped region of the *Drosophila melanogaster* X chromosome. *Genetics* 194(2):375–387.
- Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos Trans R Soc B* 372(1736):20160455.
- Stern C. 1926. An effect of temperature and age on crossing-over in the first chromosome of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 12(8):530–532.
- Stevison LS, Noor MA. 2010. Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. *J Mol Evol.* 71(5-6):332–345.
- Stuktevanat H. 1929. The genetics of *Drosophila simulans*. Washington: Carnegie Institute.
- Sturtevant AH, Beadle GW. 1936. The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. *Genetics* 21(5):554–604.
- Szauter P. 1984. An analysis of regional constraints on exchange in *Drosophila melanogaster* using recombination-defective meiotic mutants. *Genetics* 106(1):45–71.
- Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. *Cell* 33(1):25–35.
- Termolino P, Cremona G, Consiglio MF, Conicella C. 2016. Insights into epigenetic landscape of recombination-free regions. *Chromosoma* 125(2):301–308.
- Tobler R, et al. 2014. Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Mol Biol Evol.* 31(2):364–375.
- True JR, Mercer JM, Laurie CC. 1996. Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* 142(2):507–523.

Associate editor: Esther Angert